Theory Project Proposal

CSCI 470/575: Introduction to Machine Learning

Theory Project

Preference: High

Paper Title: Transfer Learning from Speaker Verification to Multispeaker Text-To-Speech Synthesis

Paper Authors: Ye Jia et al.[Yu Zhang Ron J. Weiss Quan Wang Jonathan Shen Fei Ren Zhifeng Chen Patrick Nguyen Ruoming Pang Ignacio Lopez Moreno Yonghui Wu], Google Inc.

Conference: NIPS (2018)

Research Area(s): text-speech recognition, voice cloning, audio synthesis
This paper aims to build a TTS system which can generate natural speech for a variety of speakers in a data efficient manner. Text-to-speech or Speech synthesis is the artificial production of human speech. This paper focuses on a neural network based system for text-to-speech (TTS) synthesis that is able to generate speech audio in the voice of different speakers. The implementation would be consists of three models : the encoderwhich compute the vector from speech signals, synthesizer which predicts a mel spectrogram from a sequence of grapheme or phoneme inputs, conditioned on the speaker embedding vector and vocoder which converts the spectrogram into time domain waveforms.

Data: I will be using dataset of VCTK and LibriSpeech publicly avaialable on http://www.openslr.org/12/
VCTK contains 44 hours of clean speech from 109 speakers, the majority of which have British accents. LibriSpeech consists of the union of the two clean training sets, comprising 436 hours of speech from 1,172 speakers, sampled at 16 kHz which have US accent.

Experiment: I will do my experiments on VCTK dataset by spliting the dataset into three subsets: train, validation and test. And compare the results of both the datasets. As mentioned in the paper, I will also compare the results based on follwing para,eters 1. Speech neutralness, 2. Speaker similarity, 3. Speaker verification, 4. Speaker embedding space, 5. Number of speakers, Fictiuos speakers. For working with models I will use the pretrained models available on [https://github.com/CorentinJ/Real-Time-Voice-Cloning/wiki/Pretrained-models.

Timeline

Work Breakdown Structure (WBS)

1. Theory project proposal
   1.1. Analysis of both the dataset (Pooja)
   1.2. Working with pre-trained models
2. Proposal presentation
   2.1. Implementation of VCTK and LibriSpeech (using python)
3. Progress report
   3.1. Verify data set results (Pooja)
4. Final report
   4.1. Final project report (Pooja)
5. Final Presentation
   5.1. Final Presentation (Pooja)

Critical Path The critical path of the project with expected completion dates of each task is

- 1 Task (Pooja) - (10/22/19)
- 2 Task (Pooja) - (10/29/19)
- 3 Task (Pooja) - (11/16/19)
- 4 Task (Pooja) - (11/28/19)
- 5 Task (Pooja) - (12/03/19)